This file and several accompanying files contain the solutions to the odd-numbered problems in the book *Econometric Analysis of Cross Section and Panel Data*, by Jeffrey M. Wooldridge, MIT Press, 2002. The empirical examples are solved using various versions of Stata, with some dating back to Stata 4.0. Partly out of laziness, but also because it is useful for students to see computer output, I have included Stata output in most cases rather than type tables. In some cases, I do more hand calculations than are needed in current versions of Stata.

Currently, there are some missing solutions. I will update the solutions occasionally to fill in the missing solutions, and to make corrections. For some problems I have given answers beyond what I originally asked. Please report any mistakes or discrepencies you might come across by sending me e-mail at wooldri1@msu.edu.

CHAPTER 2

2.1. a. $\dfrac{\partial \mathrm{E}(y|x_1,x_2)}{\partial x_1} = \beta_1 + \beta_4 x_2$ and $\dfrac{\partial \mathrm{E}(y|x_1,x_2)}{\partial x_2} = \beta_2 + 2\beta_3 x_2 + \beta_4 x_1$.

b. By definition, $\mathrm{E}(u|x_1,x_2) = 0$. Because x_2^2 and x_1x_2 are just functions of (x_1,x_2), it does not matter whether we also condition on them: $\mathrm{E}(u|x_1,x_2,x_2^2,x_1x_2) = 0$.

c. All we can say about $\mathrm{Var}(u|x_1,x_2)$ is that it is nonnegative for all x_1 and x_2: $\mathrm{E}(u|x_1,x_2) = 0$ in no way restricts $\mathrm{Var}(u|x_1,x_2)$.

2.3. a. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$, where u has a zero mean given x_1 and x_2: $\mathrm{E}(u|x_1,x_2) = 0$. We can say nothing further about u.

b. $\partial \mathrm{E}(y|x_1,x_2)/\partial x_1 = \beta_1 + \beta_3 x_2$. Because $\mathrm{E}(x_2) = 0$, $\beta_1 =$

$E[\partial E(y|x_1,x_2)/\partial x_1]$. Similarly, $\beta_2 = E[\partial E(y|x_1,x_2)/\partial x_2]$.

c. If x_1 and x_2 are independent with zero mean then $E(x_1x_2) = E(x_1)E(x_2) = 0$. Further, the covariance between x_1x_2 and x_1 is $E(x_1x_2\cdot x_1) = E(x_1^2x_2) = E(x_1^2)E(x_2)$ (by independence) $= 0$. A similar argument shows that the covariance between x_1x_2 and x_2 is zero. But then the linear projection of x_1x_2 onto $(1,x_1,x_2)$ is identically zero. Now just use the law of iterated projections (Property LP.5 in Appendix 2A):

$$\begin{aligned} L(y|1,x_1,x_2) &= L(\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_1x_2|1,x_1,x_2) \\ &= \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3L(x_1x_2|1,x_1,x_2) \\ &= \beta_0 + \beta_1x_1 + \beta_2x_2. \end{aligned}$$

d. Equation (2.47) is more useful because it allows us to compute the partial effects of x_1 and x_2 at *any* values of x_1 and x_2. Under the assumptions we have made, the linear projection in (2.48) does have as its slope coefficients on x_1 and x_2 the partial effects at the population average values of x_1 and x_2 -- zero in both cases -- but it does not allow us to obtain the partial effects at any other values of x_1 and x_2. Incidentally, the main conclusions of this problem go through if we allow x_1 and x_2 to have any population means.

2.5. By definition, $Var(u_1|\mathbf{x},\mathbf{z}) = Var(y|\mathbf{x},\mathbf{z})$ and $Var(u_2|\mathbf{x}) = Var(y|\mathbf{x})$. By assumption, these are constant and necessarily equal to $\sigma_1^2 \equiv Var(u_1)$ and $\sigma_2^2 \equiv Var(u_2)$, respectively. But then Property CV.4 implies that $\sigma_2^2 \geq \sigma_1^2$. This simple conclusion means that, when error variances are constant, the error variance falls as more explanatory variables are conditioned on.

2.7. Write the equation in error form as

$$y = g(\mathbf{x}) + \mathbf{z}\boldsymbol{\beta} + u, \quad \mathrm{E}(u|\mathbf{x},\mathbf{z}) = 0.$$

Take the expected value of this equation conditional only on $\mathbf{x}$:

$$\mathrm{E}(y|\mathbf{x}) = g(\mathbf{x}) + [\mathrm{E}(\mathbf{z}|\mathbf{x})]\boldsymbol{\beta},$$

and subtract this from the first equation to get

$$y - \mathrm{E}(y|\mathbf{x}) = [\mathbf{z} - \mathrm{E}(\mathbf{z}|\mathbf{x})]\boldsymbol{\beta} + u$$

or $\tilde{y} = \tilde{\mathbf{z}}\boldsymbol{\beta} + u$. Because $\tilde{\mathbf{z}}$ is a function of $(\mathbf{x},\mathbf{z})$, $\mathrm{E}(u|\tilde{\mathbf{z}}) = 0$ (since $\mathrm{E}(u|\mathbf{x},\mathbf{z}) = 0$), and so $\mathrm{E}(\tilde{y}|\tilde{\mathbf{z}}) = \tilde{\mathbf{z}}\boldsymbol{\beta}$. This basic result is fundamental in the literature on estimating *partial linear models*. First, one estimates $\mathrm{E}(y|\mathbf{x})$ and $\mathrm{E}(\mathbf{z}|\mathbf{x})$ using very flexible methods, typically, so-called *nonparametric methods*. Then, after obtaining residuals of the form $\tilde{y}_i \equiv y_i - \hat{\mathrm{E}}(y_i|\mathbf{x}_i)$ and $\tilde{\mathbf{z}}_i \equiv \mathbf{z}_i - -\hat{\mathrm{E}}(\mathbf{z}_i|\mathbf{x}_i)$, $\boldsymbol{\beta}$ is estimated from an OLS regression $\tilde{y}_i$ on $\tilde{\mathbf{z}}_i$, $i = 1,\ldots,N$. Under general conditions, this kind of nonparametric partialling-out procedure leads to a $\sqrt{N}$-consistent, asymptotically normal estimator of $\boldsymbol{\beta}$. See Robinson (1988) and Powell (1994).

CHAPTER 3

3.1. To prove Lemma 3.1, we must show that for all $\varepsilon > 0$, there exists $b_\varepsilon < \infty$ and an integer N_ε such that $\mathrm{P}[|x_N| \geq b_\varepsilon] < \varepsilon$, all $N \geq N_\varepsilon$. We use the following fact: since $x_N \xrightarrow{p} a$, for any $\varepsilon > 0$ there exists an integer N_ε such that $\mathrm{P}[|x_N - a| > 1] < \varepsilon$ for all $N \geq N_\varepsilon$. [The existence of N_ε is implied by Definition 3.3(1).] But $|x_N| = |x_N - a + a| \leq |x_N - a| + |a|$ (by the triangle inequality), and so $|x_N| - |a| \leq |x_N - a|$. It follows that $\mathrm{P}[|x_N| - |a| > 1] \leq \mathrm{P}[|x_N - a| > 1]$. Therefore, in Definition 3.3(3) we can take $b_\varepsilon \equiv |a| + 1$ (irrespective of the value of ε) and then the existence of N_ε follows from Definition 3.3(1).

3.3. This follows immediately from Lemma 3.1 because $\mathbf{g}(\mathbf{x}_N) \xrightarrow{p} \mathbf{g}(\mathbf{c})$.

3.5. a. Since $\text{Var}(\bar{y}_N) = \sigma^2/N$, $\text{Var}[\sqrt{N}(\bar{y}_N - \mu)] = N(\sigma^2/N) = \sigma^2$.

b. By the CLT, $\sqrt{N}(\bar{y}_N - \mu) \overset{a}{\sim} \text{Normal}(0,\sigma^2)$, and so $\text{Avar}[\sqrt{N}(\bar{y}_N - \mu)] = \sigma^2$.

c. We Obtain $\text{Avar}(\bar{y}_N)$ by dividing $\text{Avar}[\sqrt{N}(\bar{y}_N - \mu)]$ by N. Therefore, $\text{Avar}(\bar{y}_N) = \sigma^2/N$. As expected, this coincides with the actual variance of $\bar{y}_N$.

d. The asymptotic standard deviation of $\bar{y}_N$ is the square root of its asymptotic variance, or $\sigma/\sqrt{N}$.

e. To obtain the asymptotic standard error of $\bar{y}_N$, we need a consistent estimator of σ. Typically, the unbiased estimator of σ^2 is used: $\hat{\sigma}^2 = (N - 1)^{-1}\sum_{i=1}^{N}(y_i - \bar{y}_N)^2$, and then $\hat{\sigma}$ is the positive square root. The asymptotic standard error of $\bar{y}_N$ is simply $\hat{\sigma}/\sqrt{N}$.

3.7. a. For $\theta > 0$ the natural logarithim is a continuous function, and so $\text{plim}[\log(\hat{\theta})] = \log[\text{plim}(\hat{\theta})] = \log(\theta) = \gamma$.

b. We use the delta method to find $\text{Avar}[\sqrt{N}(\hat{\gamma} - \gamma)]$. In the scalar case, if $\hat{\gamma} = g(\hat{\theta})$ then $\text{Avar}[\sqrt{N}(\hat{\gamma} - \gamma)] = [dg(\theta)/d\theta]^2\text{Avar}[\sqrt{N}(\hat{\theta} - \theta)]$. When $g(\theta) = \log(\theta)$ -- which is, of course, continuously differentiable -- $\text{Avar}[\sqrt{N}(\hat{\gamma} - \gamma)] = (1/\theta)^2\text{Avar}[\sqrt{N}(\hat{\theta} - \theta)]$.

c. In the scalar case, the asymptotic standard error of $\hat{\gamma}$ is generally $|dg(\hat{\theta})/d\theta|\cdot\text{se}(\hat{\theta})$. Therefore, for $g(\theta) = \log(\theta)$, $\text{se}(\hat{\gamma}) = \text{se}(\hat{\theta})/\hat{\theta}$. When $\hat{\theta} = 4$ and $\text{se}(\hat{\theta}) = 2$, $\hat{\gamma} = \log(4) \approx 1.39$ and $\text{se}(\hat{\gamma}) = 1/2$.

d. The asymptotic t statistic for testing H_0: $\theta = 1$ is $(\hat{\theta} - 1)/\text{se}(\hat{\theta}) = 3/2 = 1.5$.

e. Because $\gamma = \log(\theta)$, the null of interest can also be stated as H_0: $\gamma =$

0. The t statistic based on $\hat{\gamma}$ is about 1.39/(.5) = 2.78. This leads to a very strong rejection of H_0, whereas the t statistic based on $\hat{\theta}$ is, at best, marginally significant. The lesson is that, using the Wald test, we can change the outcome of hypotheses tests by using nonlinear transformations.

3.9. By the delta method,

$$\text{Avar}[\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})] = \mathbf{G}(\boldsymbol{\theta})\mathbf{V}_1\mathbf{G}(\boldsymbol{\theta})', \ \text{Avar}[\sqrt{N}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma})] = \mathbf{G}(\boldsymbol{\theta})\mathbf{V}_2\mathbf{G}(\boldsymbol{\theta})',$$

where $\mathbf{G}(\boldsymbol{\theta}) = \nabla_{\theta}\mathbf{g}(\boldsymbol{\theta})$ is $Q \times P$. Therefore,

$$\text{Avar}[\sqrt{N}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma})] - \text{Avar}[\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})] = \mathbf{G}(\boldsymbol{\theta})(\mathbf{V}_2 - \mathbf{V}_1)\mathbf{G}(\boldsymbol{\theta})'.$$

By assumption, $\mathbf{V}_2 - \mathbf{V}_1$ is positive semi-definite, and therefore $\mathbf{G}(\boldsymbol{\theta})(\mathbf{V}_2 - \mathbf{V}_1)\mathbf{G}(\boldsymbol{\theta})'$ is p.s.d. This completes the proof.

CHAPTER 4

4.1. a. Exponentiating equation (4.49) gives

$$\begin{aligned} wage &= \exp(\beta_0 + \beta_1 married + \beta_2 educ + \mathbf{z}\boldsymbol{\gamma} + u) \\ &= \exp(u)\exp(\beta_0 + \beta_1 married + \beta_2 educ + \mathbf{z}\boldsymbol{\gamma}). \end{aligned}$$

Therefore,

$$\text{E}(wage|\mathbf{x}) = \text{E}[\exp(u)|\mathbf{x}]\exp(\beta_0 + \beta_1 married + \beta_2 educ + \mathbf{z}\boldsymbol{\gamma}),$$

where $\mathbf{x}$ denotes all explanatory variables. Now, if u and $\mathbf{x}$ are independent then $\text{E}[\exp(u)|\mathbf{x}] = \text{E}[\exp(u)] = \delta_0$, say. Therefore

$$\text{E}(wage|\mathbf{x}) = \delta_0\exp(\beta_0 + \beta_1 married + \beta_2 educ + \mathbf{z}\boldsymbol{\gamma}).$$

Now, finding the proportionate difference in this expectation at $married = 1$ and $married = 0$ (with all else equal) gives $\exp(\beta_1) - 1$; all other factors cancel out. Thus, the percentage difference is $100\cdot[\exp(\beta_1) - 1]$.

b. Since $\theta_1 = 100\cdot[\exp(\beta_1) - 1] = g(\beta_1)$, we need the derivative of g with respect to β_1: $dg/d\beta_1 = 100\cdot\exp(\beta_1)$. The asymptotic standard error of $\hat{\theta}_1$ using the delta method is obtained as the absolute value of $d\hat{g}/d\beta_1$ times $\mathrm{se}(\hat{\beta}_1)$:

$$\mathrm{se}(\hat{\theta}_1) = [100\cdot\exp(\hat{\beta}_1)]\cdot\mathrm{se}(\hat{\beta}_1).$$

c. We can evaluate the conditional expectation in part (a) at two levels of education, say $educ_0$ and $educ_1$, all else fixed. The proportionate change in expected wage from $educ_0$ to $educ_1$ is

$$[\exp(\beta_2 educ_1) - \exp(\beta_2 educ_0)]/\exp(\beta_2 educ_0)$$
$$= \exp[\beta_2(educ_1 - educ_0)] - 1 = \exp(\beta_2\Delta educ) - 1.$$

Using the same arguments in part (b), $\hat{\theta}_2 = 100\cdot[\exp(\beta_2\Delta educ) - 1]$ and

$$\mathrm{se}(\hat{\theta}_2) = 100\cdot|\Delta educ|\exp(\hat{\beta}_2\Delta educ)\mathrm{se}(\hat{\beta}_2)$$

d. For the estimated version of equation (4.29), $\hat{\beta}_1 = .199$, $\mathrm{se}(\hat{\beta}_1) = .039$, $\hat{\beta}_2 = .065$, $\mathrm{se}(\hat{\beta}_2) = .006$. Therefore, $\hat{\theta}_1 = 22.01$ and $\mathrm{se}(\hat{\theta}_1) = 4.76$. For $\hat{\theta}_2$ we set $\Delta educ = 4$. Then $\hat{\theta}_2 = 29.7$ and $\mathrm{se}(\hat{\theta}_2) = 3.11$.

4.3. a. Not in general. The conditional variance can always be written as $\mathrm{Var}(u|\mathbf{x}) = \mathrm{E}(u^2|\mathbf{x}) - [\mathrm{E}(u|\mathbf{x})]^2$; if $\mathrm{E}(u|\mathbf{x}) \neq 0$, then $\mathrm{E}(u^2|\mathbf{x}) \neq \mathrm{Var}(u|\mathbf{x})$.

b. It could be that $\mathrm{E}(\mathbf{x}'u) = \mathbf{0}$, in which case OLS is consistent, and $\mathrm{Var}(u|\mathbf{x})$ is constant. But, generally, the usual standard errors would not be valid unless $\mathrm{E}(u|\mathbf{x}) = 0$.

4.5. Write equation (4.50) as $\mathrm{E}(y|\mathbf{w}) = \mathbf{w}\boldsymbol{\delta}$, where $\mathbf{w} = (\mathbf{x}, z)$. Since $\mathrm{Var}(y|\mathbf{w}) = \sigma^2$, it follows by Theorem 4.2 that Avar $\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})$ is $\sigma^2[\mathrm{E}(\mathbf{w}'\mathbf{w})]^{-1}$, where $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}', \hat{\gamma})'$. Importantly, because $\mathrm{E}(\mathbf{x}'z) = \mathbf{0}$, $\mathrm{E}(\mathbf{w}'\mathbf{w})$ is block diagonal, with upper block $\mathrm{E}(\mathbf{x}'\mathbf{x})$ and lower block $\mathrm{E}(z^2)$. Inverting $\mathrm{E}(\mathbf{w}'\mathbf{w})$ and focusing on the upper $K \times K$ block gives

$$\text{Avar } \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sigma^2[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}.$$

Next, we need to find Avar $\sqrt{N}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$. It is helpful to write $y = \mathbf{x}\boldsymbol{\beta} + v$ where $v = \gamma z + u$ and $u \equiv y - \mathrm{E}(y|\mathbf{x},z)$. Because $\mathrm{E}(\mathbf{x}'z) = \mathbf{0}$ and $\mathrm{E}(\mathbf{x}'u) = \mathbf{0}$, $\mathrm{E}(\mathbf{x}'v) = \mathbf{0}$. Further, $\mathrm{E}(v^2|\mathbf{x}) = \gamma^2\mathrm{E}(z^2|\mathbf{x}) + \mathrm{E}(u^2|\mathbf{x}) + 2\gamma\mathrm{E}(zu|\mathbf{x}) = \gamma^2\mathrm{E}(z^2|\mathbf{x}) + \sigma^2$, where we use $\mathrm{E}(zu|\mathbf{x},z) = z\mathrm{E}(u|\mathbf{x},z) = 0$ and $\mathrm{E}(u^2|\mathbf{x},z) = \mathrm{Var}(y|\mathbf{x},z) = \sigma^2$. Unless $\mathrm{E}(z^2|\mathbf{x})$ is constant, the equation $y = \mathbf{x}\boldsymbol{\beta} + v$ generally violates the homoskedasticity assumption OLS.3. So, without further assumptions,

$$\text{Avar } \sqrt{N}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = [\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}\mathrm{E}(v^2\mathbf{x}'\mathbf{x})[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}.$$

Now we can show Avar $\sqrt{N}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ − Avar $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is positive semi-definite by writing

$$\begin{aligned}
&\text{Avar } \sqrt{N}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \text{Avar } \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= [\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}\mathrm{E}(v^2\mathbf{x}'\mathbf{x})[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1} - \sigma^2[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1} \\
&= [\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}\mathrm{E}(v^2\mathbf{x}'\mathbf{x})[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1} - \sigma^2[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}\mathrm{E}(\mathbf{x}'\mathbf{x})[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1} \\
&= [\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}[\mathrm{E}(v^2\mathbf{x}'\mathbf{x}) - \sigma^2\mathrm{E}(\mathbf{x}'\mathbf{x})][\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}.
\end{aligned}$$

Because $[\mathrm{E}(\mathbf{x}'\mathbf{x})]^{-1}$ is positive definite, it suffices to show that $\mathrm{E}(v^2\mathbf{x}'\mathbf{x}) - \sigma^2\mathrm{E}(\mathbf{x}'\mathbf{x})$ is p.s.d. To this end, let $h(\mathbf{x}) \equiv \mathrm{E}(z^2|\mathbf{x})$. Then by the law of iterated expectations, $\mathrm{E}(v^2\mathbf{x}'\mathbf{x}) = \mathrm{E}[\mathrm{E}(v^2|\mathbf{x})\mathbf{x}'\mathbf{x}] = \gamma^2\mathrm{E}[h(\mathbf{x})\mathbf{x}'\mathbf{x}] + \sigma^2\mathrm{E}(\mathbf{x}'\mathbf{x})$. Therefore, $\mathrm{E}(v^2\mathbf{x}'\mathbf{x}) - \sigma^2\mathrm{E}(\mathbf{x}'\mathbf{x}) = \gamma^2\mathrm{E}[h(\mathbf{x})\mathbf{x}'\mathbf{x}]$, which, when $\gamma \neq 0$, is actually a positive definite matrix except by fluke. In particular, if $\mathrm{E}(z^2|\mathbf{x}) = \mathrm{E}(z^2) = \eta^2 > 0$ (in which case $y = \mathbf{x}\boldsymbol{\beta} + v$ satisfies the homoskedasticity assumption OLS.3), $\mathrm{E}(v^2\mathbf{x}'\mathbf{x}) - \sigma^2\mathrm{E}(\mathbf{x}'\mathbf{x}) = \gamma^2\eta^2\mathrm{E}(\mathbf{x}'\mathbf{x})$, which is positive definite.

4.7. a. One important omitted factor in u is family income: students that come from wealthier families tend to do better in school, other things equal. Family income and PC ownership are positively correlated because the probability of owning a PC increases with family income. Another factor in u

is quality of high school. This may also be correlated with *PC*: a student who had more exposure with computers in high school may be more likely to own a computer.

b. $\hat{\beta}_3$ is *likely* to have an upward bias because of the positive correlation between u and *PC*, but it is not clear-cut because of the other explanatory variables in the equation. If we write the linear projection

$$u = \delta_0 + \delta_1 hsGPA + \delta_2 SAT + \delta_3 PC + r$$

then the bias is upward if δ_3 is greater than zero. This measures the partial correlation between u (say, family income) and *PC*, and it is likely to be positive.

c. If data on family income can be collected then it can be included in the equation. If family income is not available sometimes level of parents' education is. Another possibility is to use average house value in each student's home zip code, as zip code is often part of school records. Proxies for high school quality might be faculty-student ratios, expenditure per student, average teacher salary, and so on.

4.9. a. Just subtract $\log(y_{-1})$ from both sides:

$$\Delta\log(y) = \beta_0 + \mathbf{x}\boldsymbol{\beta} + (\alpha_1 - 1)\log(y_{-1}) + u.$$

Clearly, the intercept and slope estimates on $\mathbf{x}$ will be the same. The coefficient on $\log(y_{-1})$ changes.

b. For simplicity, let $w = \log(y)$, $w_{-1} = \log(y_{-1})$. Then the population slope coefficient in a simple regression is always $\alpha_1 = \mathrm{Cov}(w_{-1},w)/\mathrm{Var}(w_{-1})$. But, by assumption, $\mathrm{Var}(w) = \mathrm{Var}(w_{-1})$, so we can write $\alpha_1 = \mathrm{Cov}(w_{-1},w)/(\sigma_{w_{-1}}\sigma_w)$, where $\sigma_{w_{-1}} = \mathrm{sd}(w_{-1})$ and $\sigma_w = \mathrm{sd}(w)$. But $\mathrm{Corr}(w_{-1},w) = \mathrm{Cov}(w_{-1},w)/(\sigma_{w_{-1}}\sigma_w)$, and since a correlation coefficient is always between -1 and 1, the result follows.

4.11. Here is some Stata output obtained to answer this question:

```
. reg lwage exper tenure married south urban black educ iq kww

  Source |       SS       df       MS                  Number of obs =     935
---------+------------------------------               F(  9,   925) =   37.28
   Model |  44.0967944     9  4.89964382               Prob > F      =  0.0000
Residual |  121.559489   925  .131415664               R-squared     =  0.2662
---------+------------------------------               Adj R-squared =  0.2591
   Total |  165.656283   934  .177362188               Root MSE      =  .36251

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   exper |   .0127522   .0032308      3.947   0.000       .0064117    .0190927
  tenure |   .0109248   .0024457      4.467   0.000        .006125    .0157246
 married |   .1921449   .0389094      4.938   0.000       .1157839    .2685059
   south |  -.0820295   .0262222     -3.128   0.002      -.1334913   -.0305676
   urban |   .1758226   .0269095      6.534   0.000       .1230118    .2286334
   black |  -.1303995   .0399014     -3.268   0.001      -.2087073   -.0520917
    educ |   .0498375    .007262      6.863   0.000       .0355856    .0640893
      iq |   .0031183   .0010128      3.079   0.002       .0011306    .0051059
     kww |    .003826   .0018521      2.066   0.039       .0001911    .0074608
   _cons |   5.175644    .127776     40.506   0.000       4.924879    5.426408
------------------------------------------------------------------------------

. test iq kww

 ( 1)  iq = 0.0
 ( 2)  kww = 0.0

       F(  2,   925) =    8.59
            Prob > F =    0.0002
```

a. The estimated return to education using both *IQ* and *KWW* as proxies for ability is about 5%. When we used no proxy the estimated return was about 6.5%, and with only *IQ* as a proxy it was about 5.4%. Thus, we have an even lower estimated return to education, but it is still practically nontrivial and statistically very significant.

b. We can see from the *t* statistics that these variables are going to be jointly significant. The *F* test verifies this, with *p*-value = .0002.

c. The wage differential between nonblacks and blacks does not disappear. Blacks are estimated to earn about 13% less than nonblacks, holding all other factors fixed.

4.13. a. Using the 90 counties for 1987 gives

```
. reg lcrmrte lprbarr lprbconv lprbpris lavgsen if d87

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  4,    85) =   15.15
       Model |  11.1549601     4  2.78874002           Prob > F      =  0.0000
    Residual |  15.6447379    85   .18405574           R-squared     =  0.4162
-------------+------------------------------           Adj R-squared =  0.3888
       Total |   26.799698    89  .301120202           Root MSE      =  .42902

------------------------------------------------------------------------------
     lcrmrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.7239696   .1153163    -6.28   0.000    -.9532493   -.4946899
    lprbconv |  -.4725112   .0831078    -5.69   0.000    -.6377519   -.3072706
    lprbpris |   .1596698   .2064441     0.77   0.441    -.2507964     .570136
     lavgsen |   .0764213   .1634732     0.47   0.641    -.2486073    .4014499
       _cons |  -4.867922   .4315307   -11.28   0.000    -5.725921   -4.009923
------------------------------------------------------------------------------
```

Because of the log-log functional form, all coefficients are elasticities. The elasticities of crime with respect to the arrest and conviction probabilities are the sign we expect, and both are practically and statistically significant. The elasticities with respect to the probability of serving a prison term and the average sentence length are positive but are statistically insignificant.

b. To add the previous year's crime rate we first generate the lag:

```
. gen lcrmr_1 = lcrmrte[_n-1] if d87
(540 missing values generated)

. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lcrmr_1 if d87
```

```
      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  5,    84) =  113.90
       Model |  23.3549731     5  4.67099462           Prob > F      =  0.0000
    Residual |   3.4447249    84   .04100863           R-squared     =  0.8715
-------------+------------------------------           Adj R-squared =  0.8638
       Total |   26.799698    89  .301120202           Root MSE      =  .20251

------------------------------------------------------------------------------
     lcrmrte |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.1850424   .0627624    -2.95   0.004    -.3098523   -.0602325
    lprbconv |  -.0386768   .0465999    -0.83   0.409    -.1313457    .0539921
    lprbpris |  -.1266874   .0988505    -1.28   0.204    -.3232625    .0698876
     lavgsen |  -.1520228   .0782915    -1.94   0.056    -.3077141    .0036684
     lcrmr_1 |   .7798129   .0452114    17.25   0.000     .6899051    .8697208
       _cons |  -.7666256   .3130986    -2.45   0.016    -1.389257   -.1439946
------------------------------------------------------------------------------
```

There are some notable changes in the coefficients on the original variables. The elasticities with respect to *prbarr* and *prbconv* are much smaller now, but still have signs predicted by a deterrent-effect story. The conviction probability is no longer statistically significant. Adding the lagged crime rate changes the signs of the elasticities with respect to *prbpris* and *avgsen*, and the latter is almost statistically significant at the 5% level against a two-sided alternative (p-value = .056). Not surprisingly, the elasticity with respect to the lagged crime rate is large and very statistically significant. (The elasticity is also statistically different from unity.)

c. Adding the logs of the nine wage variables gives the following:

```
. reg lcrmrte lprbarr lprbconv lprbpris lavgsen lcrmr_1 lwcon-lwloc if d87

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F( 14,    75) =   43.81
       Model |  23.8798774    14  1.70570553           Prob > F      =  0.0000
    Residual |  2.91982063    75  .038930942           R-squared     =  0.8911
-------------+------------------------------           Adj R-squared =  0.8707
       Total |   26.799698    89  .301120202           Root MSE      =  .19731

------------------------------------------------------------------------------
     lcrmrte |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
    lprbarr |  -.1725122   .0659533    -2.62   0.011    -.3038978   -.0411265
   lprbconv |  -.0683639    .049728    -1.37   0.173    -.1674273    .0306994
   lprbpris |  -.2155553   .1024014    -2.11   0.039    -.4195493   -.0115614
    lavgsen |  -.1960546   .0844647    -2.32   0.023     -.364317   -.0277923
    lcrmr_1 |   .7453414   .0530331    14.05   0.000     .6396942    .8509887
      lwcon |  -.2850008   .1775178    -1.61   0.113    -.6386344    .0686327
      lwtuc |   .0641312    .134327     0.48   0.634    -.2034619    .3317244
      lwtrd |    .253707   .2317449     1.09   0.277    -.2079524    .7153665
      lwfir |  -.0835258   .1964974    -0.43   0.672    -.4749687    .3079171
      lwser |   .1127542   .0847427     1.33   0.187    -.0560619    .2815703
      lwmfg |   .0987371   .1186099     0.83   0.408    -.1375459    .3350201
      lwfed |   .3361278   .2453134     1.37   0.175    -.1525615    .8248172
      lwsta |   .0395089   .2072112     0.19   0.849    -.3732769    .4522947
      lwloc |  -.0369855   .3291546    -0.11   0.911    -.6926951     .618724
      _cons |  -3.792525   1.957472    -1.94   0.056    -7.692009    .1069592
------------------------------------------------------------------------------

. testparm lwcon-lwloc

 ( 1)  lwcon = 0.0
 ( 2)  lwtuc = 0.0
 ( 3)  lwtrd = 0.0
 ( 4)  lwfir = 0.0
 ( 5)  lwser = 0.0
 ( 6)  lwmfg = 0.0
 ( 7)  lwfed = 0.0
 ( 8)  lwsta = 0.0
 ( 9)  lwloc = 0.0

       F(  9,    75) =    1.50
            Prob > F =    0.1643
```

The nine wage variables are jointly insignificant even at the 15% level. Plus, the elasticities are not consistently positive or negative. The two largest elasticities -- which also have the largest absolute *t* statistics -- have the opposite sign. These are with respect to the wage in construction (-.285) and the wage for federal employees (.336).

d. Using the "robust" option in Stata, which is appended to the "reg" command, gives the heteroskedasiticity-robust *F* statistic as *F* = 2.19 and *p*-value = .032. (This *F* statistic is the heteroskedasticity-robust Wald statistic divided by the number of restrictions being tested, nine in this example. The division by the number of restrictions turns the asymptotic chi-square statistic into one that roughly has an F distribution.)

4.15. a. Because each x_j has finite second moment, $\text{Var}(\mathbf{x}\boldsymbol{\beta}) < \infty$. Since $\text{Var}(u) < \infty$, $\text{Cov}(\mathbf{x}\boldsymbol{\beta},u)$ is well-defined. But each x_j is uncorrelated with u, so $\text{Cov}(\mathbf{x}\boldsymbol{\beta},u) = 0$. Therefore, $\text{Var}(y) = \text{Var}(\mathbf{x}\boldsymbol{\beta}) + \text{Var}(u)$, or $\sigma_y^2 = \text{Var}(\mathbf{x}\boldsymbol{\beta}) + \sigma_u^2$.

b. This is nonsense when we view the $\mathbf{x}_i$ as random draws along with y_i. The statement "$\text{Var}(u_i) = \sigma^2 = \text{Var}(y_i)$ for all i" assumes that the regressors are nonrandom (or $\boldsymbol{\beta} = \mathbf{0}$, which is not a very interesting case). This is another example of how the assumption of nonrandom regressors can lead to counterintuitive conclusions. Suppose that an element of the error term, say z, which is uncorrelated with each x_j, suddenly becomes observed. When we add z to the regressor list, the error changes, and so does the error variance. (It gets smaller.) In the vast majority of economic applications, it makes no sense to think we have access to the entire set of factors that one would ever want to control for, so we should allow for error variances to change across different models for the same response variable.

c. Write $R^2 = 1 - \text{SSR/SST} = 1 - (\text{SSR}/N)/(\text{SST}/N)$. Therefore, $\text{plim}(R^2) = 1 - \text{plim}[(\text{SSR}/N)/(\text{SST}/N)] = 1 - [\text{plim}(\text{SSR}/N)]/[\text{plim}(\text{SST}/N)] = 1 - \sigma_u^2/\sigma_y^2 = \boldsymbol{\rho}^2$, where we use the fact that SSR/N is a consistent estimator of σ_u^2 and SST/N is a consistent estimator of σ_y^2.

d. The derivation in part (c) assumed nothing about $\text{Var}(u|\mathbf{x})$. The population R-squared depends on only the *unconditional* variances of u and y. Therefore, regardless of the nature of heteroskedasticity in $\text{Var}(u|\mathbf{x})$, the usual R-squared consistently estimates the population R-squared. Neither R-squared nor the adjusted R-squared has desirable finite-sample properties, such as unbiasedness, so the only analysis we can do in any generality involves asymptotics. The statement in the problem is simply wrong.

CHAPTER 5

5.1. Define $\mathbf{x}_1 \equiv (\mathbf{z}_1, y_2)$ and $x_2 \equiv \hat{v}_2$, and let $\hat{\boldsymbol{\beta}} \equiv (\hat{\boldsymbol{\beta}}_1', \hat{\rho}_1)'$ be OLS estimator from (5.52), where $\hat{\boldsymbol{\beta}}_1 = (\hat{\boldsymbol{\delta}}_1', \hat{\alpha}_1)'$. Using the hint, $\hat{\boldsymbol{\beta}}_1$ can also be obtained by partitioned regression:

(i) Regress $\mathbf{x}_1$ onto $\hat{v}_2$ and save the residuals, say $\ddot{\mathbf{x}}_1$.

(ii) Regress y_1 onto $\ddot{\mathbf{x}}_1$.

But when we regress $\mathbf{z}_1$ onto $\hat{v}_2$, the residuals are just $\mathbf{z}_1$ since $\hat{v}_2$ is orthogonal in sample to $\mathbf{z}$. (More precisely, $\sum_{i=1}^{N} \mathbf{z}_{i1}'\hat{v}_{i2} = \mathbf{0}$.) Further, because we can write $y_2 = \hat{y}_2 + \hat{v}_2$, where $\hat{y}_2$ and $\hat{v}_2$ are orthogonal in sample, the residuals from regressing y_2 onto $\hat{v}_2$ are simply the first stage fitted values, $\hat{y}_2$. In other words, $\ddot{\mathbf{x}}_1 = (\mathbf{z}_1, \hat{y}_2)$. But the 2SLS estimator of $\boldsymbol{\beta}_1$ is obtained exactly from the OLS regression y_1 on $\mathbf{z}_1$, $\hat{y}_2$.

5.3. a. There may be unobserved health factors correlated with smoking behavior that affect infant birth weight. For example, women who smoke during pregnancy may, on average, drink more coffee or alcohol, or eat less nutritious meals.

b. Basic economics says that *packs* should be negatively correlated with cigarette price, although the correlation might be small (especially because price is aggregated at the state level). At first glance it seems that cigarette price should be exogenous in equation (5.54), but we must be a little careful. One component of cigarette price is the state tax on cigarettes. States that have lower taxes on cigarettes may also have lower quality of health care, on average. Quality of health care is in *u*, and so maybe cigarette price fails the exogeneity requirement for an IV.

c. OLS is followed by 2SLS (IV, in this case):

```
. reg lbwght male parity lfaminc packs

  Source |       SS       df       MS                  Number of obs =    1388
---------+------------------------------               F(  4,  1383) =   12.55
   Model |  1.76664363     4  .441660908               Prob > F      =  0.0000
Residual |    48.65369  1383  .035179819               R-squared     =  0.0350
---------+------------------------------               Adj R-squared =  0.0322
   Total |  50.4203336  1387  .036352079               Root MSE      =  .18756

------------------------------------------------------------------------------
  lbwght |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |   .0262407   .0100894      2.601   0.009       .0064486    .0460328
  parity |   .0147292   .0056646      2.600   0.009       .0036171    .0258414
 lfaminc |   .0180498   .0055837      3.233   0.001       .0070964    .0290032
   packs |  -.0837281   .0171209     -4.890   0.000      -.1173139   -.0501423
   _cons |   4.675618   .0218813    213.681   0.000       4.632694    4.718542
------------------------------------------------------------------------------

. reg lbwght male parity lfaminc packs (male parity lfaminc cigprice)
                                                                      (2SLS)
  Source |       SS       df       MS                  Number of obs =    1388
---------+------------------------------               F(  4,  1383) =    2.39
   Model | -91.3500269     4 -22.8375067               Prob > F      =  0.0490
Residual |  141.770361  1383  .102509299               R-squared     =       .
---------+------------------------------               Adj R-squared =       .
   Total |  50.4203336  1387  .036352079               Root MSE      =  .32017

------------------------------------------------------------------------------
  lbwght |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   packs |   .7971063   1.086275      0.734   0.463      -1.333819    2.928031
    male |   .0298205    .017779      1.677   0.094      -.0050562    .0646972
  parity |  -.0012391   .0219322     -0.056   0.955       -.044263    .0417848
 lfaminc |    .063646   .0570128      1.116   0.264      -.0481949    .1754869
   _cons |   4.467861   .2588289     17.262   0.000       3.960122    4.975601
------------------------------------------------------------------------------
```

(Note that Stata automatically shifts endogenous explanatory variables to the beginning of the list when report coefficients, standard errors, and so on.)

The difference between OLS and IV in the estimated effect of *packs* on *bwght* is huge. With the OLS estimate, one more pack of cigarettes is estimated to reduce *bwght* by about 8.4%, and is statistically significant. The IV estimate has the opposite sign, is huge in magnitude, and is not statistically significant. The sign and size of the smoking effect are not realistic.

d. We can see the problem with IV by estimating the reduced form for *packs*:

```
. reg packs male parity lfaminc cigprice

  Source |       SS       df       MS                  Number of obs =    1388
---------+------------------------------               F(  4,  1383) =   10.86
   Model |  3.76705108     4   .94176277               Prob > F      =  0.0000
Residual |  119.929078  1383  .086716615               R-squared     =  0.0305
---------+------------------------------               Adj R-squared =  0.0276
   Total |  123.696129  1387  .089182501               Root MSE      =  .29448

------------------------------------------------------------------------------
   packs |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |  -.0047261   .0158539     -0.298   0.766      -.0358264    .0263742
  parity |   .0181491   .0088802      2.044   0.041       .0007291    .0355692
 lfaminc |  -.0526374   .0086991     -6.051   0.000      -.0697023   -.0355724
cigprice |    .000777   .0007763      1.001   0.317      -.0007459    .0022999
   _cons |   .1374075   .1040005      1.321   0.187      -.0666084    .3414234
------------------------------------------------------------------------------
```

The reduced form estimates show that *cigprice* does not significantly affect *packs*; in fact, the coefficient on *cigprice* is not the sign we expect. Thus, *cigprice* fails as an IV for *packs* because *cigprice* is not partially correlated with *packs* (with a sensible sign for the correlation). This is separate from the problem that *cigprice* may not truly be exogenous in the birth weight equation.

5.5. Under the null hypothesis that q and $\mathbf{z}_2$ are uncorrelated, $\mathbf{z}_1$ and $\mathbf{z}_2$ are exogenous in (5.55) because each is uncorrelated with u_1. Unfortunately, y_2

is correlated with u_1, and so the regression of y_1 on $\mathbf{z}_1$, y_2, $\mathbf{z}_2$ does not produce a consistent estimator of $\mathbf{0}$ on $\mathbf{z}_2$ even when $\mathrm{E}(\mathbf{z}_2'q) = \mathbf{0}$. We could find that $\hat{\boldsymbol{\psi}}_1$ from this regression is statistically different from zero even when q and $\mathbf{z}_2$ are uncorrelated -- in which case we would incorrectly conclude that $\mathbf{z}_2$ is not a valid IV candidate. Or, we might fail to reject H_0: $\boldsymbol{\psi}_1 = \mathbf{0}$ when $\mathbf{z}_2$ and q are correlated -- in which case we incorrectly conclude that the elements in $\mathbf{z}_2$ are valid as instruments.

The point of this exercise is that one cannot simply add instrumental variable candidates in the structural equation and then test for significance of these variables using OLS. This is the sense in which identification cannot be tested. With a single endogenous variable, we must take a stand that at least one element of $\mathbf{z}_2$ is uncorrelated with q.

5.7. a. If we plug $q = (1/\delta_1)q_1 - (1/\delta_1)a_1$ into equation (5.45) we get

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \eta_1 q_1 + v - \eta_1 a_1, \qquad (5.56)$$

where $\eta_1 \equiv (1/\delta_1)$. Now, since the z_h are redundant in (5.45), they are uncorrelated with the structural error, v (by definition of redundancy). Further, we have assumed that the z_h are uncorrelated with a_1. Since each x_j is also uncorrelated with $v - \eta_1 a_1$, we can estimate (5.56) by 2SLS using instruments $(1,x_1,\ldots,x_K,z_1,z_2,\ldots,z_M)$ to get consistent of the β_j and η_1.

Given all of the zero correlation assumptions, what we need for identification is that at least one of the z_h appears in the reduced form for q_1. More formally, in the linear projection

$$q_1 = \pi_0 + \pi_1 x_1 + \ldots + \pi_K x_K + \pi_{K+1} z_1 + \ldots + \pi_{K+M} z_M + r_1,$$

at least one of π_{K+1}, ..., π_{K+M} must be different from zero.

b. We need family background variables to be redundant in the log(*wage*) equation once ability (and other factors, such as *educ* and *exper*), have been controlled for. The idea here is that family background may influence ability but should have no partial effect on log(*wage*) once ability has been accounted for. For the rank condition to hold, we need family background variables to be correlated with the indicator, q_1, say *IQ*, once the x_j have been netted out. This is likely to be true if we think that family background and ability are (partially) correlated.

c. Applying the procedure to the data set in NLS80.RAW gives the following results:

```
. reg lwage exper tenure educ married south urban black iq (exper tenure educ
married south urban black meduc feduc sibs)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     722
-------------+------------------------------           F(  8,   713) =   25.81
       Model |  19.6029198     8  2.45036497           Prob > F      =  0.0000
    Residual |  107.208996   713  .150363248           R-squared     =  0.1546
-------------+------------------------------           Adj R-squared =  0.1451
       Total |  126.811916   721  .175883378           Root MSE      =  .38777

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .0154368   .0077077      2.00   0.046      .0003044    .0305692
      tenure |   .0076754   .0030956      2.48   0.013      .0015979    .0137529
        educ |   .0161809   .0261982      0.62   0.537      -.035254    .0676158
     married |   .1901012   .0467592      4.07   0.000      .0982991    .2819033
       south |   -.047992   .0367425     -1.31   0.192     -.1201284    .0241444
       urban |   .1869376   .0327986      5.70   0.000      .1225442    .2513311
       black |   .0400269   .1138678      0.35   0.725     -.1835294    .2635832
       exper |   .0162185   .0040076      4.05   0.000      .0083503    .0240867
       _cons |   4.471616    .468913      9.54   0.000         3.551    5.392231
------------------------------------------------------------------------------

. reg lwage exper tenure educ married south urban black kww (exper tenure educ
married south urban black meduc feduc sibs)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     722
-------------+------------------------------           F(  8,   713) =   25.70
```

	SS	df	MS			
Model	19.820304	8	2.477538	Prob > F	=	0.0000
Residual	106.991612	713	.150058361	R-squared	=	0.1563
				Adj R-squared	=	0.1468
Total	126.811916	721	.175883378	Root MSE	=	.38737

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
kww	.0249441	.0150576	1.66	0.098	-.0046184	.0545067
tenure	.0051145	.0037739	1.36	0.176	-.0022947	.0125238
educ	.0260808	.0255051	1.02	0.307	-.0239933	.0761549
married	.1605273	.0529759	3.03	0.003	.0565198	.2645347
south	-.091887	.0322147	-2.85	0.004	-.1551341	-.0286399
urban	.1484003	.0411598	3.61	0.000	.0675914	.2292093
black	-.0424452	.0893695	-0.47	0.635	-.2179041	.1330137
exper	.0068682	.0067471	1.02	0.309	-.0063783	.0201147
_cons	5.217818	.1627592	32.06	0.000	4.898273	5.537362

Even though there are 935 men in the sample, only 722 are used for the estimation, because data are missing on *meduc* and *feduc*. What we could do is define binary indicators for whether the corresponding variable is missing, set the missing values to zero, and then use the binary indicators as instruments along with *meduc*, *feduc*, and *sibs*. This would allow us to use all 935 observations.

The return to education is estimated to be small and insignificant whether *IQ* or *KWW* used is used as the indicator. This could be because family background variables do not satisfy the appropriate redundancy condition, or they might be correlated with a_1. (In both first-stage regressions, the F statistic for joint significance of *meduc*, *feduc*, and *sibs* have p-values below .002, so it seems the family background variables are sufficiently partially correlated with the ability indicators.)

5.9. Define $\theta_4 = \beta_4 - \beta_3$, so that $\beta_4 = \beta_3 + \theta_4$. Plugging this expression into the equation and rearranging gives

$$\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3(twoyr + fouryr) + \theta_4 fouryr + u$$
$$= \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3\ totcoll + \theta_4 fouryr + u,$$

where $totcoll = twoyr + fouryr$. Now, just estimate the latter equation by 2SLS using $exper$, $exper^2$, $dist2yr$ and $dist4yr$ as the full set of instruments. We can use the t statistic on $\hat{\theta}_4$ to test H_0: $\theta_4 = 0$ against H_1: $\theta_4 > 0$.

5.11. Following the hint, let y_2^0 be the linear projection of y_2 on $\mathbf{z}_2$, let a_2 be the projection error, and assume that $\boldsymbol{\lambda}_2$ is known. (The results on generated regressors in Section 6.1.1 show that the argument carries over to the case when $\boldsymbol{\lambda}_2$ is estimated.) Plugging in $y_2 = y_2^0 + a_2$ gives

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2^0 + \alpha_1 a_2 + u_1.$$

Effectively, we regress y_1 on $\mathbf{z}_1$, y_2^0. The key consistency condition is that each explanatory is orthogonal to the composite error, $\alpha_1 a_2 + u_1$. By assumption, $E(\mathbf{z}'u_1) = \mathbf{0}$. Further, $E(y_2^0 a_2) = 0$ by construction. The problem is that $E(\mathbf{z}_1' a_2) \neq \mathbf{0}$ necessarily because $\mathbf{z}_1$ was not included in the linear projection for y_2. Therefore, OLS will be inconsistent for all parameters in general. Contrast this with 2SLS when y_2^* is the projection on $\mathbf{z}_1$ and $\mathbf{z}_2$: $y_2 = y_2^* + r_2 = \mathbf{z}\boldsymbol{\pi}_2 + r_2$, where $E(\mathbf{z}'r_2) = \mathbf{0}$. The second step regression (assuming that $\boldsymbol{\pi}_2$ is known) is essentially

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2^* + \alpha_1 r_2 + u_1.$$

Now, r_2 is uncorrelated with $\mathbf{z}$, and so $E(\mathbf{z}_1' r_2) = \mathbf{0}$ and $E(y_2^* r_2) = 0$. The lesson is that one must be very careful if manually carrying out 2SLS by explicitly doing the first- and second-stage regressions.

5.13. a. In a simple regression model with a single IV, the IV estimate of the slope can be written as $\hat{\beta}_1 = \left(\sum_{i=1}^{N}(z_i - \bar{z})(y_i - \bar{y})\right)/\left(\sum_{i=1}^{N}(z_i - \bar{z})(x_i - \bar{x})\right) =$

$\left(\sum_{i=1}^{N} z_i(y_i - \bar{y})\right) / \left(\sum_{i=1}^{N} z_i(x_i - \bar{x})\right)$. Now the numerator can be written as

$$\sum_{i=1}^{N} z_i(y_i - \bar{y}) = \sum_{i=1}^{N} z_i y_i - \left(\sum_{i=1}^{N} z_i\right)\bar{y} = N_1\bar{y}_1 - N_1\bar{y} = N_1(\bar{y}_1 - \bar{y}).$$

where $N_1 = \sum_{i=1}^{N} z_i$ is the number of observations in the sample with $z_i = 1$ and $\bar{y}_1$ is the average of the y_i over the observations with $z_i = 1$. Next, write $\bar{y}$ as a weighted average: $\bar{y} = (N_0/N)\bar{y}_0 + (N_1/N)\bar{y}_1$, where the notation should be clear. Straightforward algebra shows that $\bar{y}_1 - \bar{y} = [(N - N_1)/N]\bar{y}_1 - (N_0/N)\bar{y}_0 = (N_0/N)(\bar{y}_1 - \bar{y}_0)$. So the numerator of the IV estimate is $(N_0 N_1/N)(\bar{y}_1 - \bar{y}_0)$. The same argument shows that the denominator is $(N_0 N_1/N)(\bar{x}_1 - \bar{x}_0)$. Taking the ratio proves the result.

b. If x is also binary -- representing some "treatment" -- $\bar{x}_1$ is the fraction of observations receiving treatment when $z_i = 1$ and $\bar{x}_0$ is the fraction receiving treatment when $z_i = 0$. So, suppose $x_i = 1$ if person i participates in a job training program, and let $z_i = 1$ if person i is eligible for participation in the program. Then $\bar{x}_1$ is the fraction of people participating in the program out of those made eligibile, and $\bar{x}_0$ is the fraction of people participating who are not eligible. (When eligibility is necessary for participation, $\bar{x}_0 = 0$.) Generally, $\bar{x}_1 - \bar{x}_0$ is the difference in participation rates when $z = 1$ and $z = 0$. So the difference in the mean response between the $z = 1$ and $z = 0$ groups gets divided by the difference in participation rates across the two groups.

5.15. In $\mathrm{L}(\mathbf{x}|\mathbf{z}) = \mathbf{z}\boldsymbol{\Pi}$, we can write $\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\Pi}_{11} & \mathbf{0} \\ \boldsymbol{\Pi}_{12} & \mathbf{I}_{K_2} \end{pmatrix}$, where $\mathbf{I}_{K_2}$ is the K_2 x K_2 identity matrix, $\mathbf{0}$ is the L_1 x K_2 zero matrix, $\boldsymbol{\Pi}_{11}$ is L_1 x K_1, and $\boldsymbol{\Pi}_{12}$ is K_2 x K_1. As in Problem 5.12, the rank condition holds if and only if $\mathrm{rank}(\boldsymbol{\Pi}) = K$.

a. If for some x_j, the vector $\mathbf{z}_1$ does not appear in $\mathrm{L}(x_j|\mathbf{z})$, then $\boldsymbol{\Pi}_{11}$ has a column which is entirely zeros. But then that column of $\mathbf{\Pi}$ can be written as a linear combination of the last K_2 elements of $\mathbf{\Pi}$, which means rank($\mathbf{\Pi}$) < K. Therefore, a necessary condition for the rank condition is that no columns of $\mathbf{\Pi}_{11}$ be exactly zero, which means that at least one z_h must appear in the reduced form of each x_j, $j = 1,\ldots,K_1$.

b. Suppose $K_1 = 2$ and $L_1 = 2$, where z_1 appears in the reduced form form both x_1 and x_2, but z_2 appears in neither reduced form. Then the 2 x 2 matrix $\mathbf{\Pi}_{11}$ has zeros in its second row, which means that the second row of $\mathbf{\Pi}$ is all zeros. It cannot have rank K, in that case. Intuitively, while we began with two instruments, only one of them turned out to be partially correlated with x_1 and x_2.

c. Without loss of generality, we assume that z_j appears in the reduced form for x_j; we can simply reorder the elements of $\mathbf{z}_1$ to ensure this is the case. Then $\mathbf{\Pi}_{11}$ is a K_1 x K_1 diagonal matrix with nonzero diagonal elements. Looking at $\mathbf{\Pi} = \begin{pmatrix} \mathbf{\Pi}_{11} & \mathbf{0} \\ \mathbf{\Pi}_{12} & \mathbf{I}_{K_2} \end{pmatrix}$, we see that if $\mathbf{\Pi}_{11}$ is diagonal with all nonzero diagonals then $\mathbf{\Pi}$ is lower triangular with all nonzero diagonal elements. Therefore, rank $\mathbf{\Pi} = K$.

CHAPTER 6

6.1. a. Here is abbreviated Stata output for testing the null hypothesis that *educ* is exogenous:

```
. qui reg educ nearc4 nearc2 exper expersq black south smsa reg661-reg668
smsa66

. predict v2hat, resid
```

```
. reg lwage educ exper expersq black south smsa reg661-reg668 smsa66 v2hat

------------------------------------------------------------------------------
   lwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .1570594   .0482814      3.253   0.001        .0623912    .2517275
   exper |   .1188149   .0209423      5.673   0.000        .0777521    .1598776
 expersq |  -.0023565   .0003191     -7.384   0.000       -.0029822   -.0017308
   black |  -.1232778   .0478882     -2.574   0.010       -.2171749   -.0293807
   south |  -.1431945   .0261202     -5.482   0.000       -.1944098   -.0919791
    smsa |    .100753   .0289435      3.481   0.000        .0440018    .1575042
  reg661 |   -.102976   .0398738     -2.583   0.010       -.1811588   -.0247932
  reg662 |  -.0002286   .0310325     -0.007   0.994       -.0610759    .0606186
  reg663 |   .0469556   .0299809      1.566   0.117       -.0118296    .1057408
  reg664 |  -.0554084   .0359807     -1.540   0.124       -.1259578    .0151411
  reg665 |   .0515041   .0436804      1.179   0.238       -.0341426    .1371509
  reg666 |   .0699968   .0489487      1.430   0.153       -.0259797    .1659733
  reg667 |   .0390596   .0456842      0.855   0.393        -.050516    .1286352
  reg668 |  -.1980371   .0482417     -4.105   0.000       -.2926273   -.1034468
  smsa66 |   .0150626   .0205106      0.734   0.463       -.0251538    .0552789
   v2hat |  -.0828005   .0484086     -1.710   0.087        -.177718    .0121169
   _cons |   3.339687    .821434      4.066   0.000        1.729054    4.950319
------------------------------------------------------------------------------
```

The t statistic on $\hat{v}_2$ is -1.71, which is not significant at the 5% level against a two-sided alternative. The negative correlation between u_1 and *educ* is essentially the same finding that the 2SLS estimated return to education is larger than the OLS estimate. In any case, I would call this marginal evidence that *educ* is endogenous. (Depending on the application or purpose of a study, the same researcher may take t = -1.71 as evidence for or against endogeneity.)

b. To test the single overidentifying restiction we obtain the 2SLS residuals:

```
. qui reg lwage educ exper expersq black south smsa reg661-reg668 smsa66
(nearc4 nearc2 exper expersq black south smsa reg661-reg668 smsa66)

. predict uhat1, resid
```

Now, we regress the 2SLS residuals on all exogenous variables:

```
. reg uhat1 exper expersq black south smsa reg661-reg668 smsa66 nearc4 nearc2

  Source |       SS       df       MS                  Number of obs =    3010
```

```
---------+------------------------------           F( 16,  2993) =    0.08
   Model |  .203922832    16  .012745177           Prob > F      =  1.0000
Residual |  491.568721  2993  .164239466           R-squared     =  0.0004
---------+------------------------------           Adj R-squared = -0.0049
   Total |  491.772644  3009  .163433913           Root MSE      =  .40526
```

The test statistic is the sample size times the *R*-squared from this regression:

```
. di 3010*.0004
1.204

. di chiprob(1,1.2)
.27332168
```

The *p*-value, obtained from a χ_1^2 distribution, is about .273, so the instruments pass the overidentification test.

6.3. a. We need prices to satisfy two requirements. First, *calories* and *protein* must be partially correlated with prices of food. While this is easy to test for each by estimating the two reduced forms, the rank condition could still be violated (although see Problem 15.5c). In addition, we must also assume prices are exogenous in the productivity equation. Ideally, prices vary because of things like transportation costs that are not systematically related to regional variations in individual productivity. A potential problem is that prices reflect food quality and that features of the food other than calories and protein appear in the disturbance u_1.

b. Since there are two endogenous explanatory variables we need at least two prices.

c. We would first estimate the two reduced forms for *calories* and *protein* by regressing each on a constant, *exper*, $exper^2$, *educ*, and the *M* prices, p_1, ..., p_M. We obtain the residuals, $\hat{v}_{21}$ and $\hat{v}_{22}$. Then we would run the regression log(*produc*) on 1, *exper*, $exper^2$, *educ*, $\hat{v}_{21}$, $\hat{v}_{22}$ and do a joint significance test on $\hat{v}_{21}$ and $\hat{v}_{22}$. We could use a standard F test or use a heteroskedasticity-robust test.

6.5. a. For simplicity, absorb the intercept in $\mathbf{x}$, so $y = \mathbf{x}\boldsymbol{\beta} + u$, $\mathrm{E}(u|\mathbf{x}) = 0$, $\mathrm{Var}(u|\mathbf{x}) = \sigma^2$. In these tests, $\hat{\sigma}^2$ is implictly SSR/N -- there is no degrees of freedom adjustment. (In any case, the df adjustment makes no difference asymptotically.) So $\hat{u}_i^2 - \hat{\sigma}^2$ has a zero sample average, which means that

$$N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\hat{u}_i^2 - \hat{\sigma}^2) = N^{-1/2}\sum_{i=1}^{N}\mathbf{h}_i'(\hat{u}_i^2 - \hat{\sigma}^2).$$

Next, $N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)' = O_p(1)$ by the central limit theorem and $\hat{\sigma}^2 - \sigma^2 = o_p(1)$. So $N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\hat{\sigma}^2 - \sigma^2) = O_p(1)\cdot o_p(1) = o_p(1)$. Therefore, so far we have

$$N^{-1/2}\sum_{i=1}^{N}\mathbf{h}_i'(\hat{u}_i^2 - \hat{\sigma}^2) = N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\hat{u}_i^2 - \sigma^2) + o_p(1).$$

We are done with this part if we show $N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'\hat{u}_i^2 = N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'u_i^2 + o_p(1)$. Now, as in Problem 4.4, we can write $\hat{u}_i^2 = u_i^2 - 2u_i\mathbf{x}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + [\mathbf{x}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]^2$, so

$$\begin{aligned} N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'\hat{u}_i^2 &= N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'u_i^2 \\ &\quad - 2\left(N^{-1/2}\sum_{i=1}^{N}u_i(\mathbf{h}_i - \boldsymbol{\mu}_h)'\mathbf{x}_i\right)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \qquad (6.40) \\ &\quad + \left(N^{-1/2}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{x}_i \otimes \mathbf{x}_i)\right)\{\mathrm{vec}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']\}, \end{aligned}$$

where the expression for the third term follows from $[\mathbf{x}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]^2 = \mathbf{x}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{x}_i' = (\mathbf{x}_i \otimes \mathbf{x}_i)\mathrm{vec}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']$. Dropping the "-2" the second term can be written as $\left(N^{-1}\sum_{i=1}^{N}u_i(\mathbf{h}_i - \boldsymbol{\mu}_h)'\mathbf{x}_i\right)\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_p(1)\cdot O_p(1)$ because $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_p(1)$ and, under $\mathrm{E}(u_i|\mathbf{x}_i) = 0$, $\mathrm{E}[u_i(\mathbf{h}_i - \boldsymbol{\mu}_h)'\mathbf{x}_i] = \mathbf{0}$; the law of large numbers implies that the sample average is $o_p(1)$. The third term can be written as $N^{-1/2}\left(N^{-1}\sum_{i=1}^{N}(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{x}_i \otimes \mathbf{x}_i)\right)\{\mathrm{vec}[\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']\} = N^{-1/2}\cdot O_p(1)\cdot O_p(1)$, where we again use the fact that sample averages are $O_p(1)$ by the law of large numbers and $\mathrm{vec}[\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] = O_p(1)$. We have shown that the last two terms in (6.40) are $o_p(1)$, which proves part (a).

b. By part (a), the asymptotic variance of $N^{-1/2}\sum_{i=1}^{N}\mathbf{h}_i'(\hat{u}_i^2 - \hat{\sigma}^2)$ is $\mathrm{Var}[(\mathbf{h}_i - \boldsymbol{\mu}_h)'(u_i^2 - \sigma^2)] = \mathrm{E}[(u_i^2 - \sigma^2)^2(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{h}_i - \boldsymbol{\mu}_h)]$. Now $(u_i^2 - \sigma^2)^2 = u_i^4 - 2u_i^2\sigma^2 + \sigma^4$. Under the null, $\mathrm{E}(u_i^2|\mathbf{x}_i) = \mathrm{Var}(u_i|\mathbf{x}_i) = \sigma^2$ [since $\mathrm{E}(u_i|\mathbf{x}_i) = 0$ is assumed] and therefore, when we add (6.27), $\mathrm{E}[(u_i^2 - \sigma^2)^2|\mathbf{x}_i] = \kappa^2 - \sigma^4 \equiv \eta^2$. A standard iterated expectations argument gives $\mathrm{E}[(u_i^2 - \sigma^2)^2(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{h}_i - \boldsymbol{\mu}_h)] = \mathrm{E}\{\mathrm{E}[(u_i^2 - \sigma^2)^2(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{h}_i - \boldsymbol{\mu}_h)]|\mathbf{x}_i\} = \mathrm{E}\{\mathrm{E}[(u_i^2 - \sigma^2)^2|\mathbf{x}_i](\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{h}_i - \boldsymbol{\mu}_h)\}$ [since $\mathbf{h}_i = \mathbf{h}(\mathbf{x}_i)$] $= \eta^2\mathrm{E}[(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{h}_i - \boldsymbol{\mu}_h)]$. This is what we wanted to show. (Whether we do the argument for a random draw i or for random variables representing the population is a matter of taste.)

c. From part (b) and Lemma 3.8, the following statistic has an asymptotic χ_Q^2 distribution:

$$\left(N^{-1/2}\sum_{i=1}^{N}(\hat{u}_i^2 - \hat{\sigma}^2)\mathbf{h}_i\right)\{\eta^2\mathrm{E}[(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{h}_i - \boldsymbol{\mu}_h)]\}^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{h}_i'(\hat{u}_i^2 - \hat{\sigma}^2)\right).$$

Using again the fact that $\sum_{i=1}^{N}(\hat{u}_i^2 - \hat{\sigma}^2) = 0$, we can replace $\mathbf{h}_i$ with $\mathbf{h}_i - \bar{\mathbf{h}}$ in the two vectors forming the quadratic form. Then, again by Lemma 3.8, we can replace the matrix in the quadratic form with a consistent estimator, which is

$$\hat{\eta}^2\left(N^{-1}\sum_{i=1}^{N}(\mathbf{h}_i - \bar{\mathbf{h}})'(\mathbf{h}_i - \bar{\mathbf{h}})\right),$$

where $\hat{\eta}^2 = N^{-1}\sum_{i=1}^{N}(\hat{u}_i^2 - \hat{\sigma}^2)^2$. The computable statistic, after simple algebra, can be written as

$$\left(\sum_{i=1}^{N}(\hat{u}_i^2 - \hat{\sigma}^2)(\mathbf{h}_i - \bar{\mathbf{h}})\right)\left(\sum_{i=1}^{N}(\mathbf{h}_i - \bar{\mathbf{h}})'(\mathbf{h}_i - \bar{\mathbf{h}})\right)^{-1}\left(\sum_{i=1}^{N}(\mathbf{h}_i - \bar{\mathbf{h}})'(\hat{u}_i^2 - \hat{\sigma}^2)\right)/\hat{\eta}^2.$$

Now $\hat{\eta}^2$ is just the total sum of squares in the $\hat{u}_i^2$, divided by N. The numerator of the statistic is simply the explained sum of squares from the regression $\hat{u}_i^2$ on 1, $\mathbf{h}_i$, $i = 1,\ldots,N$. Therefore, the test statistic is N times the usual (centered) R-squared from the regression $\hat{u}_i^2$ on 1, $\mathbf{h}_i$, $i = 1,\ldots,N$, or NR_c^2.

d. Without assumption (6.37) we need to estimate $\mathrm{E}[(u_i^2 - \sigma^2)^2(\mathbf{h}_i - \boldsymbol{\mu}_h)'(\mathbf{h}_i - \boldsymbol{\mu}_h)]$ generally. Hopefully, the approach is by now pretty clear. We replace the population expected value with the sample average and replace any unknown parameters -- $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\mu}_h$ in this case -- with their consistent estimators (under H_0). So a generally consistent estimator of $\text{Avar}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{h}_i'(\hat{u}_i^2 - \hat{\sigma}^2)\right)$ is

$$N^{-1}\sum_{i=1}^{N}(\hat{u}_i^2 - \hat{\sigma}^2)^2(\mathbf{h}_i - \bar{\mathbf{h}})'(\mathbf{h}_i - \bar{\mathbf{h}}),$$

and the test statistic robust to heterokurtosis can be written as

$$\left(\sum_{i=1}^{N}(\hat{u}_i^2 - \hat{\sigma}^2)(\mathbf{h}_i - \bar{\mathbf{h}})\right)\left(\sum_{i=1}^{N}(\hat{u}_i^2 - \hat{\sigma}^2)^2(\mathbf{h}_i - \bar{\mathbf{h}})'(\mathbf{h}_i - \bar{\mathbf{h}})\right)^{-1} \cdot \left(\sum_{i=1}^{N}(\mathbf{h}_i - \bar{\mathbf{h}})'(\hat{u}_i^2 - \hat{\sigma}^2)\right),$$

which is easily seen to be the explained sum of squares from the regression of 1 on $(\hat{u}_i^2 - \hat{\sigma}^2)(\mathbf{h}_i - \bar{\mathbf{h}})$, $i = 1,\ldots,N$ (without an intercept). Since the total sum of squares, without demeaning, is $N = (1 + 1 + \ldots + 1)$ (N times), the statistic is equivalent to $N - \text{SSR}_0$, where SSR_0 is the sum of squared residuals.

6.7. a. The simple regression results are

```
. reg lprice ldist if y81

  Source |       SS       df       MS                  Number of obs =     142
---------+------------------------------               F(  1,   140) =   30.79
   Model |  3.86426989     1  3.86426989               Prob > F      =  0.0000
Residual |  17.5730845   140  .125522032               R-squared     =  0.1803
---------+------------------------------               Adj R-squared =  0.1744
   Total |  21.4373543   141  .152037974               Root MSE      =  .35429

------------------------------------------------------------------------------
  lprice |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   ldist |   .3648752   .0657613      5.548   0.000       .2348615    .4948889
   _cons |   8.047158   .6462419     12.452   0.000       6.769503    9.324813
------------------------------------------------------------------------------
```

This regression suggests a strong link between housing price and distance from the incinerator (as distance increases, so does housing price). The elasticity is .365 and the *t* statistic is 5.55. However, this is not a good causal regression: the incinerator may have been put near homes with lower values to begin with. If so, we would expect the positive relationship found in the simple regression even if the new incinerator had no effect on housing prices.

b. The parameter δ_3 should be positive: after the incinerator is built a house should be worth more the farther it is from the incinerator. Here is my Stata session:

```
. gen y81ldist = y81*ldist

. reg lprice y81 ldist y81ldist

  Source |       SS       df       MS                  Number of obs =     321
---------+------------------------------               F(  3,   317) =   69.22
   Model |  24.3172548     3  8.10575159               Prob > F      =  0.0000
Residual |  37.1217306   317  .117103251               R-squared     =  0.3958
---------+------------------------------               Adj R-squared =  0.3901
   Total |  61.4389853   320  .191996829               Root MSE      =   .3422

------------------------------------------------------------------------------
  lprice |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     y81 |  -.0113101   .8050622     -0.014   0.989       -1.59525     1.57263
   ldist |    .316689   .0515323      6.145   0.000       .2153006    .4180775
y81ldist |   .0481862   .0817929      0.589   0.556      -.1127394    .2091117
   _cons |   8.058468   .5084358     15.850   0.000       7.058133    9.058803
------------------------------------------------------------------------------
```

The coefficient on *ldist* reveals the shortcoming of the regression in part (a). This coefficient measures the relationship between *lprice* and *ldist* in 1978, before the incinerator was even being rumored. The effect of the incinerator is given by the coefficient on the interaction, *y81ldist*. While the direction of the effect is as expected, it is not especially large, and it is statistically insignificant anyway. Therefore, at this point, we cannot reject the null hypothesis that building the incinerator had no effect on housing prices.

c. Adding the variables listed in the problem gives

```
. reg lprice y81 ldist y81ldist lintst lintstsq larea lland age agesq rooms
baths

  Source |       SS       df       MS                  Number of obs =     321
---------+------------------------------               F( 11,   309) =  108.04
   Model |  48.7611143    11  4.43282858               Prob > F      =  0.0000
Residual |   12.677871   309  .041028709               R-squared     =  0.7937
---------+------------------------------               Adj R-squared =  0.7863
   Total |  61.4389853   320  .191996829               Root MSE      =  .20256

------------------------------------------------------------------------------
  lprice |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     y81 |   -.229847   .4877198     -0.471   0.638      -1.189519    .7298249
   ldist |   .0866424   .0517205      1.675   0.095      -.0151265    .1884113
y81ldist |   .0617759   .0495705      1.246   0.214      -.0357625    .1593143
  lintst |   .9633332   .3262647      2.953   0.003       .3213518    1.605315
lintstsq |  -.0591504   .0187723     -3.151   0.002       -.096088   -.0222128
   larea |   .3548562   .0512328      6.926   0.000       .2540468    .4556655
   lland |    .109999   .0248165      4.432   0.000       .0611683    .1588297
     age |  -.0073939   .0014108     -5.241   0.000      -.0101699   -.0046178
   agesq |   .0000315   8.69e-06      3.627   0.000       .0000144    .0000486
   rooms |   .0469214   .0171015      2.744   0.006       .0132713    .0805715
   baths |   .0958867    .027479      3.489   0.000        .041817    .1499564
   _cons |   2.305525   1.774032      1.300   0.195      -1.185185    5.796236
------------------------------------------------------------------------------
```

The incinerator effect is now larger (the elasticity is about .062) and the *t* statistic is larger, but the interaction is still statistically insignificant. Using these models and this two years of data we must conclude the evidence that housing prices were adversely affected by the new incinerator is somewhat weak.

6.9. a. The Stata results are

```
. reg ldurat afchnge highearn afhigh male married head-construc if ky

      Source |       SS       df       MS              Number of obs =    5349
-------------+------------------------------           F( 14,  5334) =   16.37
       Model |  358.441793    14  25.6029852           Prob > F      =  0.0000
    Residual |  8341.41206  5334  1.56381928           R-squared     =  0.0412
-------------+------------------------------           Adj R-squared =  0.0387
```

```
       Total |  8699.85385  5348  1.62674904           Root MSE      =  1.2505

------------------------------------------------------------------------------
      ldurat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     afchnge |   .0106274   .0449167     0.24   0.813    -.0774276    .0986824
    highearn |   .1757598   .0517462     3.40   0.001     .0743161    .2772035
      afhigh |   .2308768   .0695248     3.32   0.001     .0945798    .3671738
        male |  -.0979407   .0445498    -2.20   0.028    -.1852766   -.0106049
     married |   .1220995   .0391228     3.12   0.002     .0454027    .1987962
        head |  -.5139003   .1292776    -3.98   0.000    -.7673372   -.2604634
        neck |   .2699126   .1614899     1.67   0.095    -.0466737    .5864988
      upextr |   -.178539   .1011794    -1.76   0.078     -.376892    .0198141
       trunk |   .1264514   .1090163     1.16   0.246    -.0872651     .340168
     lowback |  -.0085967   .1015267    -0.08   0.933    -.2076305    .1904371
     lowextr |  -.1202911   .1023262    -1.18   0.240    -.3208922    .0803101
      occdis |   .2727118    .210769     1.29   0.196    -.1404816    .6859052
       manuf |  -.1606709   .0409038    -3.93   0.000    -.2408591   -.0804827
    construc |   .1101967   .0518063     2.13   0.033     .0086352    .2117581
       _cons |   1.245922   .1061677    11.74   0.000      1.03779    1.454054
------------------------------------------------------------------------------
```

The estimated coefficient on the interaction term is actually higher now, and even more statistically significant than in equation (6.33). Adding the other explanatory variables only slightly increased the standard error on the interaction term.

b. The small *R*-squared, on the order of 4.1%, or 3.9% if we used the adjusted *R*-squared, means that we cannot explain much of the variation in time on workers compensation using the variables included in the regression. This is often the case in the social sciences: it is very difficult to include the multitude of factors that can affect something like *durat*. The low *R*-squared means that making predictions of log(*durat*) would be very difficult given the factors we have included in the regression: the variation in the unobservables pretty much swamps the explained variation. However, the low *R*-squared does not mean we have a biased or consistent estimator of the effect of the policy change. Provided the Kentucky change is a good natural experiment, the OLS estimator is consistent. With over 5,000 observations, we can get a reasonably precise estimate of the effect, although the 95% confidence interval is pretty wide.

c. Using the data for Michigan to estimate the simple model gives

```
. reg ldurat afchnge highearn afhigh if mi

      Source |       SS       df       MS              Number of obs =    1524
-------------+------------------------------           F(  3,  1520) =    6.05
       Model |  34.3850177     3  11.4616726           Prob > F      =  0.0004
    Residual |  2879.96981  1520  1.89471698           R-squared     =  0.0118
-------------+------------------------------           Adj R-squared =  0.0098
       Total |  2914.35483  1523  1.91356194           Root MSE      =  1.3765

------------------------------------------------------------------------------
      ldurat |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     afchnge |   .0973808   .0847879     1.15   0.251    -.0689329    .2636945
    highearn |   .1691388   .1055676     1.60   0.109    -.0379348    .3762124
      afhigh |   .1919906   .1541699     1.25   0.213    -.1104176    .4943988
       _cons |   1.412737   .0567172    24.91   0.000     1.301485    1.523989
------------------------------------------------------------------------------
```

The coefficient on the interaction term, .192, is remarkably similar to that for Kentucky. Unfortunately, because of the many fewer observations, the *t* statistic is insignificant at the 10% level against a one-sided alternative. Asymptotic theory predicts that the standard error for Michigan will be about $(5,626/1,524)^{1/2} \approx 1.92$ larger than that for Kentucky. In fact, the ratio of standard errors is about 2.23. The difference in the KY and MI cases shows the importance of a large sample size for this kind of policy analysis.

6.11. The following is Stata output that I will use to answer the first three parts:

```
. reg lwage y85 educ y85educ exper expersq union female y85fem

      Source |       SS       df       MS              Number of obs =    1084
-------------+------------------------------           F(  8,  1075) =   99.80
       Model |  135.992074     8  16.9990092           Prob > F      =  0.0000
```

```
    Residual |  183.099094  1075  .170324738           R-squared     =  0.4262
-------------+------------------------------           Adj R-squared =  0.4219
       Total |  319.091167  1083   .29463635           Root MSE      =   .4127

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         y85 |   .1178062   .1237817     0.95   0.341     -.125075    .3606874
        educ |   .0747209   .0066764    11.19   0.000     .0616206    .0878212
     y85educ |   .0184605   .0093542     1.97   0.049      .000106     .036815
       exper |   .0295843   .0035673     8.29   0.000     .0225846     .036584
     expersq |  -.0003994   .0000775    -5.15   0.000    -.0005516   -.0002473
       union |   .2021319   .0302945     6.67   0.000     .1426888    .2615749
      female |  -.3167086   .0366215    -8.65   0.000    -.3885663    -.244851
      y85fem |    .085052    .051309     1.66   0.098    -.0156251     .185729
       _cons |   .4589329   .0934485     4.91   0.000     .2755707     .642295
------------------------------------------------------------------------------
```

a. The return to another year of education increased by about .0185, or 1.85 percentage points, between 1978 and 1985. The *t* statistic is 1.97, which is marginally significant at the 5% level against a two-sided alternative.

b. The coefficient on *y85fem* is positive and shows that the estimated gender gap declined by about 8.5 percentage points. But the *t* statistic is only significant at about the 10% level against a two-sided alternative. Still, this is suggestive of some closing of wage differentials between men and women at given levels of education and workforce experience.

c. Only the coefficient on *y85* changes if wages are measured in 1978 dollars. In fact, you can check that when 1978 wages are used, the coefficient on *y85* becomes about –.383, which shows a significant fall in real wages for given productivity characteristics and gender over the seven-year period. (But see part e for the proper interpretation of the coefficient.)

d. To answer this question, I just took the squared OLS residuals and regressed those on the year dummy, *y85*. The coefficient is about .042 with a standard error of about .022, which gives a *t* statistic of about 1.91. So there is some evidence that the variance of the unexplained part of log wages (or log real wages) has increased over time.

e. As the equation is written in the problem, the coefficient δ_0 is the growth in nominal wages for a male with no years of education! For a male with 12 years of education, we want $\theta_0 \equiv \delta_0 + 12\delta_1$. A simple way to obtain the standard error of $\hat{\theta}_0 = \hat{\delta}_0 + 12\hat{\delta}_1$ is to replace *y85·educ* with *y85·*(*educ* - 12). Simple algebra shows that, in the new model, θ_0 is the coefficient on *educ*. In Stata we have

```
. gen y85educ0 = y85*(educ - 12)

. reg lwage y85 educ y85educ0 exper expersq union female y85fem

      Source |       SS       df       MS              Number of obs =    1084
-------------+------------------------------           F(  8,  1075) =   99.80
       Model |  135.992074     8  16.9990092           Prob > F      =  0.0000
    Residual |  183.099094  1075  .170324738           R-squared     =  0.4262
-------------+------------------------------           Adj R-squared =  0.4219
       Total |  319.091167  1083   .29463635           Root MSE      =   .4127

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         y85 |   .3393326   .0340099     9.98   0.000     .2725993    .4060659
        educ |   .0747209   .0066764    11.19   0.000     .0616206    .0878212
    y85educ0 |   .0184605   .0093542     1.97   0.049      .000106     .036815
       exper |   .0295843   .0035673     8.29   0.000     .0225846     .036584
     expersq |  -.0003994   .0000775    -5.15   0.000    -.0005516   -.0002473
       union |   .2021319   .0302945     6.67   0.000     .1426888    .2615749
      female |  -.3167086   .0366215    -8.65   0.000    -.3885663   -.244851
      y85fem |    .085052    .051309     1.66   0.098    -.0156251    .185729
       _cons |   .4589329   .0934485     4.91   0.000     .2755707     .642295
------------------------------------------------------------------------------
```

So the growth in nominal wages for a man with *educ* = 12 is about .339, or 33.9%. [We could use the more accurate estimate, obtained from exp(.339) -1.] The 95% confidence interval goes from about 27.3 to 40.6.